Controlled, distributed data management of an Antarctic time series

Adam Leadbetter[1], David Connor[2], Nathan Cunningham[2], and Sarah Reynolds[1]

[1]British Oceanographic Data Centre, Joseph Proudman Building, 6 Brownlow Street, Liverpool, L3 5DA, United Kingdom
[2]British Antarctic Survey, High Cross, Madingley Road, Cambridge, CB3 0ET, United Kingdom
alead@bodc.ac.uk

Introduction

The Rothera Time Series (RaTS) is a complex, long-term dataset. Sampling began in January 1997 at a site approximately 4km offshore of the British Antarctic Survey station at Rothera. Weather and sea ice conditions permitting, an upper ocean hydrography cast and a water bottle sample (at 15m depth) are taken every fifth day in summer and every seventh day in winter. Time series of currents at depth have also been recorded from moorings deployed by the BAS research vessel. Some 30 scientific publications use the RaTS data[a].

Data Rescue

The RaTS dataset was considered to be 'at risk' as it was held in a 'final' form only on a BAS scientist's laptop, with the data therefore in danger of being lost. Recognising the importance of these data and that well managed data allows the progress of science, steps were taken to secure the data and make them available to the wider community. This data rescue exercise was facilitated by new data management techniques and the growing relationship between BAS and the British Oceanographic Data Centre (BODC)[b].

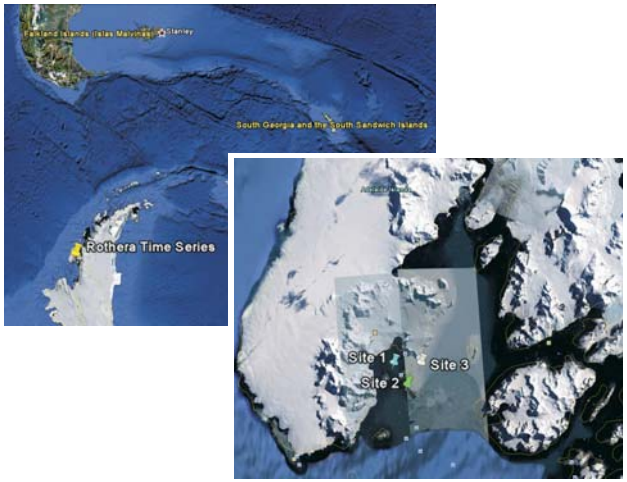

Figure 1. Geographical location of the Rothera Time Series. Site 1 is the primary sampling location. Sites 2 & 3 are used when site 1 and site 2 are respectively inaccessible. Images from Google Earth.

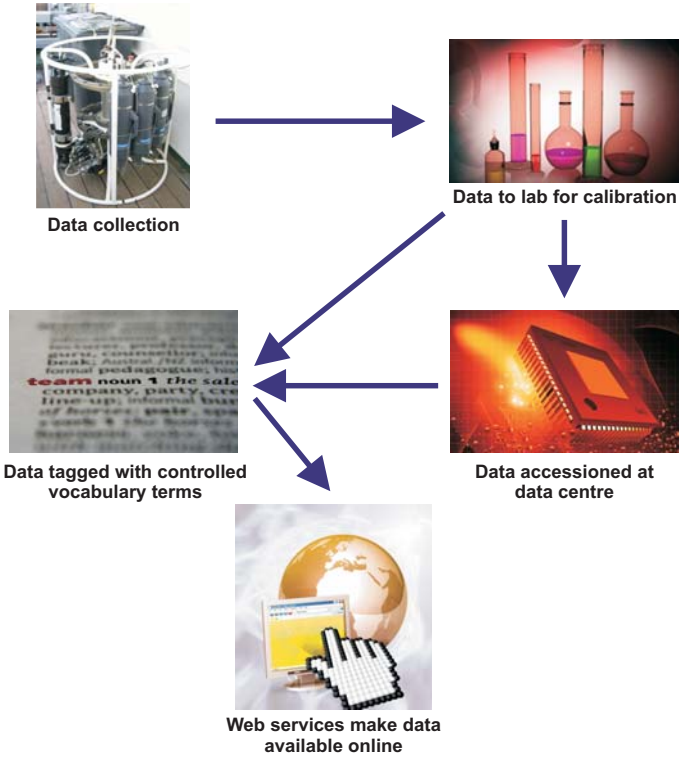

Figure 2. The flow of the RaTS data: from its collection in the field to being available for search and download on the world wide web.

Method

The RaTS data are collected in the field and then transmitted back to BAS, Cambridge. Once the series data (hydrography casts, current meter time series) have been calibrated (approximately annually) they are delivered to BODC. At BODC the data are tagged with parameter usage vocabulary terms from the NERC Data Grid (NDG) controlled vocabularies, while the discrete sample data are tagged from the same vocabulary at BAS. As this vocabulary uses semantic links, tagging provides the data with semantic richness allowing it to immediately be searched using crosswalks between GCMD, ISO, MEDIN, Inspire and NDG / SeaDataNet terminology. Taxonomic data are also automatically linked to WoRMS and ITIS.

The series data have been subject to quality control measures at BODC, where all data are checked by eye for spurious points. Metadata and data documentation concering the RaTS data have been compiled to provide provenance and contextual information to sit alongside the data. These metadata ensure the completeness of the dataset.

As the NDG vocabulary allows cross-site data compatibility, BODC provide an Open Geospatial Consortium Web Feature Service to BAS which allows the full catalogue of RaTS series and discrete sample data to be searched from one web site, despite the data themselves being managed at separate physical locations.

Future Directions

The techniques used in marking up and making available the RaTS data could be extended to provide a whole-NERC portal, opening up the tools required to produce Earth System Science. The NDG vocabularies are being linked to other thesaurii through the EU NETMAR programme, deepening the semantic richness associated with the RaTS data. This dataset should also be linked with a data journal or data publishing house.

Conclusions

This work has secured, and possibly saved from loss or obsolesence, a unique dataset. As well as being secured, the dataset is now complete with a rich metadata layer, and is publicly available through BODC allowing it to be open to public scruitiny. This last point being particularly important in the current political climate.

References

a. Rothera Oceanographic and Biological Time Series (RaTS) web page. http://www.antarctica.ac.uk/rats Accessed 7th April 2010

b. Leadbetter, A.M. and Cunningham N. J. (2009). Short- and long-term management of Antarctic oceanographic data. NERC Data Management Workshop.

c. http://vocab.ndg.nerc.ac.uk

NATURAL ENVIRONMENT RESEARCH COUNCIL

BODC

British Antarctic Survey
NATURAL ENVIRONMENT RESEARCH COUNCIL